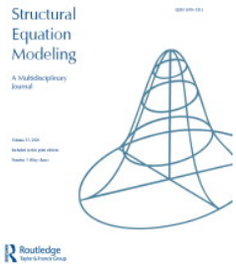
Routledge
Taylor & Francis Group



Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: www.tandfonline.com/journals/hsem20

Meaningful Latent Class Analysis: Essential Statistical Frameworks and Best Practices for the Applied Researcher

Johan Lyrvall, Jennifer Oser & Zsuzsa Bakk

To cite this article: Johan Lyrvall, Jennifer Oser & Zsuzsa Bakk (17 Jul 2026): Meaningful Latent Class Analysis: Essential Statistical Frameworks and Best Practices for the Applied Researcher, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2026.2686838

To link to this article: https://doi.org/10.1080/10705511.2026.2686838

Published online: 17 Jul 2026.

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2026.2686838

Meaningful Latent Class Analysis: Essential Statistical Frameworks and Best Practices for the Applied Researcher

Johan Lyrvall[a], Jennifer Oser[b], and Zsuzsa Bakk[c]

[a]University of Catania; [b]Ben-Gurion University of the Negev; [c]Leiden University

ABSTRACT

This article integrates existing theoretical and applied teaching resources to provide an updated overview of best practices for latent class analysis (LCA), focusing on applied researchers. We first clarify the unique usefulness of LCA compared to related latent variable models. We then focus on three fundamental research questions relevant for the applied researcher: (1) *do I have data with a possible clustering structure that can be analyzed using LCA?* (2) *what optimal number of LCs describe the possible clustering structure in my data?* and (3) *can the identified LCs in my data be predicted based on external covariates?* We conclude with summary guidance regarding modeling extensions for more complex LCA topics that are highly relevant for applied researchers, namely differential item functioning (DIF) and multilevel modeling. The topics are illustrated through empirical examples, with provision of replication code. An overview of software options for LCA is provided.

KEYWORDS

Differential item functioning; latent class analysis; model selection; multilevel modeling; software; stepwise estimation

1. Introduction

Introduced by Lazarsfeld and Henry in 1968, latent class analysis (LCA) is used to identify distinct sub-groups of individuals based on a set of observed behaviors. The LC model takes these observed behaviors to be indicators of an underlying structure of mutually exclusive latent classes, and describes the representative pattern of behaviors and proportion in the population of each latent class. For example, in economics, Angelini and Lyrvall (2025) applied LCA to identify distinct norms on governmental punitive action against public institutions supplying scandalous cultural goods based on four categorical survey items; in substance abuse research, Bray et al. (2023) applied LCA to identify distinct types of opioid users based on five categorical indicators from patient data collected in a clinical trial; in political science, Oser et al. (2023) applied LCA to identify citizenship norms based on twelve categorical survey items, and Bertsou and Caramani (2022) applied LCA to identify distinct technocratic attitudes based on twelve categorical survey items.

LCA is an instance of latent variable modeling (LVM). In the social sciences, LVM is used to identify complex constructs that cannot be measured directly, by means of indirectly measuring them via multiple observed indicators. The most common LVM approach is factor analysis, which identifies a continuous latent variable based on continuous indicators. The other LVM approaches are item-response theory, in which the latent variable is continuous and the indicators categorical, and latent profile analysis, in which the latent variable is categorical and the indicators continuous. What is unique about LCA is that it enables applied researchers to identify categorical latent variables based on categorical indicators. This is of substantial interest to scholars aiming to analyze distinct sub-groups of individuals based on data typically observed in surveys like fixed-choice survey (for example *yes*/*no* or *agree*/*neither agree nor disagree*/*disagree*).

Multivariate categorical data can also be analyzed using other methods than LCA. The most common alternative is linear scales. In linear scales, the measurement of interest is the sum of the responses of interest across the observed variables, like the total number of *yes* responses across a set of *yes*/*no* survey items. The primary benefit of LCA compared to linear scales is the ability to identify qualitative differences in response patterns that would not be revealed by a sum of responses. As an illustrative example, consider a hypothetical analysis of wine and cheese consumption based on four binary consumption measures (*consumes*/*does not consume*) for high-end wine, low-end wine, high-end cheese, and low-end cheese, with consumers clustered into two unobserved consumer segments: the luxury type selecting high-end wine and high-end cheese, and the budget type selecting low-end wine and low-end cheese. In linear scales, both the luxury type and the budget type will be represented by a quantity of 2, since both consume a total of two product categories. In LCA, the distinct consumer segments and their qualitative differences will be revealed. As such, when the research question involves identifying distinct sub-groups of individuals, the actor-oriented LCA approach has great advantages over the variable-centric linear scales approach.

In the last decades, LCA has become an increasingly popular applied methodology in the behavioral sciences. In

CONTACT Jennifer Oser oser@post.bgu.ac.il Department of Politics and Government, Ben-Gurion University of the Negev, Beer Sheva, Israel.

meeting this increasing interest, the LCA literature has in recent years put forth methodological overviews of standard LCA.[1] These tend to focus either on the general modeling framework (Nylund-Gibson & Choi, 2018; Oberski, 2016; Van Lissa et al., 2024; Weller et al., 2020) or on the more specific modeling context of LCA with covariates (Bakk & Kuha, 2021; Kim et al., 2016; Vermunt, 2025). However, while these are relevant and useful overviews, as teaching resources they are asymmetrically focused between theoretical concepts and practical model-building decisions, resulting in a lack of coherence between theoretical and pragmatic concerns in the meaningful application of LCA. In this paper, we put forth an alternative teaching resource for applied LCA, aiming to provide a balanced exposition of essential statistical framework and best practices.

The exposition is organized around three fundamental research questions in applied LCA. Specifically, we pose research questions regarding the LCA concept and data requirements, model selection for the number of LCs, and regression for LCs, in a standard LCA modeling framework. After presenting essential statistical frameworks and best practices for answering each of these fundamental research questions, we illustrate how they can be answered by means of an empirical analysis of patterns of political participation and their association with political placement on the left-right scale. Subsequently, we briefly focus on two more advanced topics: differential item functioning and multilevel models. Next, we provide a brief overview of the main software options for estimating all the discussed models. We hope that this teaching resource will support the meaningful application of LCA, clarifying the combination of technical skill and theoretical judgment required at several stages of applying LCA.

2. Research Questions

2.1. RQ 1: Do I Have Data with a Possible Clustering Structure That Can be Analyzed Using LCA?

Latent class analysis (LCA) is used to identify distinct subgroups of the population, or *latent classes*, modeled in terms of an underlying discrete variable. In applied research, the LC variable is typically defined as an unobserved construct establishing a typology of individuals. The substantive narrative around applied LCA often interprets the LC variable either based on the sub-groups or based on the concept underlying the sub-group partitioning. For simplicity, in this paper, we focus on a common interpretation based on typologies of individuals, and leave it to the reader to take this as primarily focused on the individuals or on the underlying theoretical construct. For example, in applied economics, Angelini and Lyrvall (2025) identified a typology of punitive action from the government against public institutions supplying scandalous cultural goods. More specifically, the authors identified four distinct underlying norms, which they labeled "strict punishers," "conditional punishers," "moderate free-expression advocates," and "strong free-expression advocates." For instance, the "strict punishers" tend to strongly support government use of censorship and funding cuts for public museums and libraries supplying scandalous artworks and books, whereas the "conditional punishers" tend to oppose censorship and support funding cuts.

Consider a population of N individuals $i = 1, ..., N$ which are partitioned into types based on the LC variable X, which has T categories, or LCs, $t = 1, ..., T$. In applied LCA, the number of types T is typically unknown a priori. Model selection for T is presented in the next section. In this section, we take T as given. We denote by $P(X = t)$ the marginal probability of belonging to class t. In what follows, we interpret this as the unconditional proportion of type t. In standard LCA, it is assumed that i belongs to exactly one class, so that $\sum_{t=1}^{T} P(X = t) = 1$, and that each class has at least one member, so that $P(X = t) > 0 \forall t$.

How is the LC variable X identified in the LCA methodological framework? The LC variable X is assumed to affect a set of observed categorical *indicators*. Based on the assumed relationship between X and the indicators, which we introduce in the next paragraph, the indicators are used to derive X; that is, X can be analyzed indirectly via the indicators. In applied research, the indicators are usually survey items. For example, in their applied LCA, Angelini and Lyrvall (2025) use as indicators 4 survey items measuring individuals' opinions about public-sector supply of scandalous art. Let H be the number of indicators $h = 1, ..., H$, and Y_h the h-th indicator, with categories $r_h = 1, ..., R_h$, of which i endorses exactly one. For example, Y_h can be a survey item asking the respondent to choose one of the $R_h = 4$ response options *(1) strongly agree*, *(2) mildly agree*, *(3) mildly disagree*, and *(4) strongly disagree*. We denote by $P(Y_1 = r_1, ..., Y_H = r_H)$ the marginal joint probability of a particular response pattern; by $P(Y_1 = r_1, ..., Y_H = r_H | X = t)$ the corresponding conditional joint probability for members in t; and by $P(Y_h = r_h | X = t)$ the conditional probability of a particular response on a particular indicator for members in t.

The LC model takes the marginal probability $P(Y_1 = r_1, ..., Y_H = r_H)$ to be a mixture of the T different conditional, type-specific probabilities $P(Y_1 = r_1, ..., Y_H = r_H | X = t)$, each weighted by the corresponding type proportion $P(X = t)$. We can write this *mixture assumption* as

$$P(Y_1 = r_1, ..., Y_H = r_H) = \sum_{t=1}^{T} P(X = t)$$
$$P(Y_1 = r_1, ..., Y_H = r_H | X = t).$$

As can be seen, $P(Y_1 = r_1, ..., Y_H = r_H)$ can be interpreted as the weighted average of the corresponding proportions across all the types. The standard LC model also takes the indicators to be conditionally independent of each other

[1] We focus here on literature that is similar to the present contribution, that is, academic journal articles focused only on LCA as traditionally defined (both the latent variable and its observed indicators being categorical and analyzed cross-sectionally). Additional noteworthy contributions that are beyond this scope of comparison include academic journal articles like Steinley and Brusco (2011) and Moore et al. (2025), book chapters like Steinley (2023) and Masyn (2013), and books like Collins and Lanza (2013) and Hagenaars and McCutcheon (2002).

given X. We can write this *local independence assumption* as

$$P(Y_1 = r_1, ..., Y_H = r_H|X = t) = \prod_{h=1}^{H}\prod_{r_h=1}^{R_h} P(Y_h = r_h|X = t)^{I(Y_h=r_h)},$$

where $I(Y_h = r_h)$ is equal to 1 if r_h is the observed value on Y_h, and else 0. This augmenting variable serves to include in the multiplication only the response probabilities for the observed categories on the indicators. Taking the mixture assumption and the local independence assumption together, we obtain the standard LC model equation

$$P(Y_1 = r_1, ..., Y_H = r_H) = \sum_{t=1}^{T} P(X = t) \prod_{h=1}^{H}\prod_{r_h=1}^{R_h} P(Y_h = r_h|X = t)^{I(Y_h=r_h)}. \quad (1)$$

The LC model can be separated into 2 conceptually distinct components: the *measurement model* and the *structural model*. The measurement model defines the types. This is described by the set of parameters $P(Y_h = r_h|X = t)$. The structural model describes how common each type is. This is described by the set of parameters $P(X = t)$.

Figure 1 graphically illustrates the standard LC model by means of a path diagram. The arrows illustrate that the observed values on the indicators depend on what type the respondent belongs to. The absence of arrows between the indicators illustrate their local independence.

In this paper, we do not focus on details of estimation of LC models. Here we briefly mention that LC models are typically estimated by means of maximum likelihood[2] using the expectation maximization, or EM, algorithm (Dempster et al., 1977). When using this approach, it is generally recommended to utilize multiple starts and performing the estimation multiple times to increase the probability of the fitted model being a global maximum and not a local maximum. Sometimes, although less frequently, estimation is carried out by means of a Bayesian approach involving Gibbs sampling. For details of the implementation of the EM algorithm and a Gibbs sampling algorithm for fitting the model we refer the interested reader to White and Murphy (2014). While the authors present these routines for the case of binary indicators, the routines can be straightforwardly extended to the case of polytomous indicators.

To illustrate the LC model parameters and their interpretation in an applied context, we analyze data from the 1987 General Social Survey (McCutcheon, 1987). These data include 1713 respondents surveyed as a cross-section of the American public. In this example, we identify a typology of freedom of speech norms based on 3 binary survey items. These indicators measure whether or not the respondent believes that anti-religionists should be allowed to speak; that anti-religionists should be allowed to teach; and that anti-religious books should be removed from the library. The relative frequency of the response patterns are reported in Table 1, where 1 corresponds to support for freedom of speech (allow speaking, allow teaching, do not remove books) and 0 opposition to freedom of speech (ban speaking, ban teaching, remove books). For example, we can see in the bottom row that 41% of respondents believe that anti-religionists should be allowed to speak, that anti-religionists should be allowed to teach, and that anti-religious books should be kept in the library.

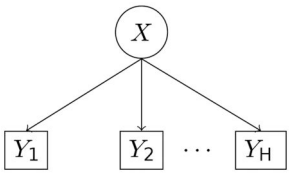


Figure 1. Path diagram for the standard LC model.

Table 1. Sample proportion of response patterns for the 1987 General Social survey freedom of speech items.

speak	*teach*	*remove*	prop.
0	0	0	0.21
0	0	1	0.07
0	1	0	0.01
1	0	0	0.08
1	1	0	0.04
1	0	1	0.16
0	1	1	0.02
1	1	1	0.41

Based on these data, we fit the model with $T = 2$ types. The resulting estimates are reported in Table 2. We present the conditional response probabilities by means of the probabilities for the answer option corresponding to support for free speech (measured by value 1 in Table 1). Class 1 is the largest, representing 62% of individuals. This type has high probabilities of supporting each form of freedom of speech: 96% probability of responding that anti-religionists should be allowed to teach, 74% probability of responding that anti-religionists should be allowed to teach, and 92% probability of responding that anti-religious books should be kept in the library. Class 2 has low probabilities of endorsing freedom of speech (or, equivalently, high probabilities of opposing freedom of speech). To see the relationship between the sample proportions of the response patterns and the estimates for the class proportions and the conditional response probabilities, consider the probability of endorsing all three items. As shown in Table 1, in the sample, this probability is 0.41. The fitted LC model shown in Table 2 replicated this probability when we plug its estimates into Equation (1): $0.62 * (0.96 * 0.74 * 0.92) + 0.38 * (0.23 * 0.04 * 0.24) = 0.41$. This empirical example illustrates how LCA can be applied to provide detailed insights into the structure of a population of interest with respect to unobserved distinct types of individuals.

2.2. RQ 2: What Optimal Number of LCs Describe the Possible Clustering Structure in my Data?

In applied LCA, the number of types T is typically an unknown quantity to be identified. Model selection for the

[2]This is performed on the basis of the log-likelihood function $\ell = \sum_{i=1}^{N} \log\left[\sum_{t=1}^{T} P(X = t) \prod_{h=1}^{H}\prod_{r_h=1}^{R_h} P(Y_h = r_{ih}|X = t)^{I(Y_h=r_{ih})}\right]$, where r_{ih} is the observed response for unit i on indicator Y_h.

Table 2. Fitted two-class LC model based on data from the 1987 General Social survey.

	Class 1	Class 2
Proportion	0.62	0.38
Response probability		
speak	0.96	0.23
teach	0.74	0.04
book	0.92	0.24

number of types involves comparing specifications with a range of different candidates (for example, 1–4). The maximum number of types that can be statistically identified depends on the data. Given the data, models with more types (greater T) have fewer degrees of freedom. In Appendix A we provide details on how to compute a model's degrees of freedom. Here we mention the key point that a model must have non-negative degrees of freedom to be statistically identifiable.

The minimum T to be included in the range of specifications is, according to applied LCA best practices, 1. The methodological appeal of considering the one-class specification lies in its conceptual contradiction. Conceptually, a sub-group structure consisting of a single sub-group cannot reasonably be considered a sub-group structure to begin with. This is because the "sub"-group is not nested within the population. Rather, the "sub"-group is the population itself. Including the one-class specification in the range of candidate Ts is equivalent to testing a hypothesis of the existence of the LC structure in the data, and the legitimacy of LCA as a quantitative strategy for the study, against a null hypothesis of the absence of the LC structure in the data. Model selection resulting in the one-class specification suggests that there are no LCs in the data, and model selection resulting in a specification with $T \geq 2$ suggests that the data do have a LC structure and that LCA is a legitimate quantitative strategy for the study.

The one-class model is always statistically identifiable. Its parameter estimates can be computed based on basic descriptive statistics. The estimates for the measurement model parameters are given by the sample proportion of response q_h to indicator Y_h, or $P(Y_h = q_h|X = t) = P(Y_h = q_h) = \frac{1}{N}\sum_{i=1}^{N} P(Y_{ih} = q_h)$. The structural model does not need to be estimated nor computed, as, per definition, it is necessarily given by $P(X = 1) = 1$.

There are different strategies for selecting the largest number of LCs to consider in the range of candidates when substantive theory does not offer a guide. One strategy is to consider a wide range of statistically identifiable specifications. A different strategy is to incrementally add one additional LC while the one or two largest candidates are selected over the specification for the preceding candidate. As an example of the third strategy, model selection may begin by estimating the one-class and two-class specifications, and observing that the latter is optimal. Then, the three-class specification may be estimated and observed not to be optimal compared to the two-class specification. Now the two-class specification can be selected as locally optimal, or, if the applied researcher is more conservative, the two-class solution may be required also to be preferable to the four-class specification (and, if not, require the four-class specification to be preferable to the five- and six-class specifications, and so on).

On what criteria is model selection typically based? Different applied LCAs tend to use different sets of selection criteria. For conciseness, we here bring popular selection criteria to the reader's attention without providing formulae. The most popular family of selection criteria is information criteria. Conceptually, information criteria aim to identify an optimal tradeoff between (maximizing) model fit and (minimizing) model complexity. The most commonly used is the Bayesian information criterion (BIC). The BIC is often used in combination with the Akaike information criterion (AIC). The rationale for their combined use is that the BIC tends to underestimate T, and AIC tends to overestimate T; yet, it should be noted, the same rationale sometimes leads researchers to using the AIC3 (Bozdogan, 1993), which tends to yield a quantity between the AIC and the BIC and perform better than these criteria when evaluated separately (see Lukočienė et al., 2010, and citations therein). Other popular selection criteria are likelihood ratio tests (LRTs) and class separation measures. LRTs are used to test the statistical significance of a candidate T to an alternative candidate $T - 1$. In LCA, the most common LRTs include the standard LRT and the bootstrap LRT, or BLRT (e.g., Dziak et al., 2014).

Class separation can be understood as the distinctiveness of the identified types. However, the appropriateness of using class separation measures for model selection is not uniformly recognized. The practice is primarily popular among applied researchers, but not widespread among methodologists. Strong class separation is sometimes taken as an indication that the identified LCs are true types, or, equivalently, that one LC is not an outlier pattern for a second LC identifying a true type. The most commonly used class separation measures are the entropy-R^2 (to be maximized toward 1) and the expected proportion of misclassification (to be minimized toward 0). Information criteria and class separation measures are conceptually combined into the ICL-BIC (Biernacki et al., 2000), an adjustment of the BIC with the integrated complete likelihood (ICL; Biernacki et al., 2010) to reward class separation, aiming to identify an optimal tradeoff between (maximizing) model fit, (minimizing) model complexity, and (maximizing) class separation. We refer the interested reader to Masyn (2013) for other, yet less common, quantitative selection criteria, including the correct model probability, Bayes factor, average posterior class membership probabilities, odds of correct classification, and odds ratios. Beyond quantitative selection criteria, applied researchers often use the substantive interpretation of the estimated models to exclude substantively meaningless class solutions from consideration. For example, in the LCA applied literature, Oser et al. (2013) used the BIC, the likelihood-ratio test, the expected proportion of misclassification, and the substantive significance of the class solutions.

Model selection is a prominent and disputed area in LCA scholarship. There is currently no universally accepted

recommendation regarding selection criteria. Deciding the selection criteria on which to base model selection among those we have reviewed is ultimately up to the discretion of the applied researcher, in consultation with published research in the relevant applied disciplines. While the present paper is not mainly focused on model selection, we highlight that methodological guidance on model selection constitutes a promising avenue for future scholarly contributions. Guidance on ideal joint use of specific model selection criteria is particularly warranted.

2.3. RQ 3: Can the Identified LCs in my Data be Predicted Based on External Covariates?

Applied LCA typically involves analyzing the association between the types and some external predictors, or *covariates*.[3] Let $\mathbf{Z} = (1, Z_1, ..., Z_P)'$ be a set of P covariates, taken to be predictors of X, and a constant 1 to allow for an intercept. A standard assumption in the LCA methodological and applied literature, and in open-source and proprietary LCA software, is that the relationship between X and $\mathbf{Z}$ is logistic; multinomial logistic when $T \geq 3$, and binomial logistic when $T = 2$.

We emphasize the conceptual distinction between the indicators $Y_1, ..., Y_H$ and the covariates $Z_1, ..., Z_P$. The indicators serve as measurements of the unobserved LC variable X, while the covariates affect the structure of X in the population. Consider, for example, the applied LCA of freedom of speech norms from Sub-section 2.1, with the addition of two hypothetical covariates: sex and education. In this context, the sex and education of individual i is taken to be realized first. Second, given the probability of adhering to each norm for an individual with i's sex and education, the type of i is realized. Third, given i's type, i is taken to respond to the survey, realizing the response pattern.

The standard extension of Equation (1) to include the covariates is such that $\mathbf{Z}$ enters only into the structural model. We thus obtain

$$P(Y_1 = r_1, ..., Y_H = r_H|\mathbf{Z}) = \sum_{t=1}^{T} P(X = t|\mathbf{Z}) \prod_{h=1}^{H} \prod_{r_h=1}^{R_h} P(Y_h = r_h|X = t)^{I(Y_h = r_h)}. \quad (2)$$

Figure 2 graphically illustrates this model by means of a path diagram, showing that what type the respondent belongs to depends on the covariates, and that the covariates have a direct effect only on the types, not on the indicators. Because of the assumption of a logistic relationship between X and $\mathbf{Z}$, we can write the structural model as

$$P(X = t|\mathbf{Z}) = \frac{\exp(\boldsymbol{\gamma}_t'\mathbf{Z})}{1 + \sum_{s=2}^{T} \exp(\boldsymbol{\gamma}_s'\mathbf{Z})}, \quad (3)$$

where $\boldsymbol{\gamma}_t = (\gamma_{t0}, \gamma_{t1}, ..., \gamma_{tP})'$ is the set logistic coefficients, consisting of an intercept γ_{t0} and P logistic slope coefficients

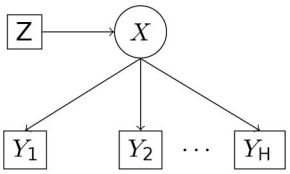


Figure 2. Path diagram for the LC model with covariates.

corresponding to $Z_1, ..., Z_P$. This general parameterization allows for a multinomial logistic definition when $T \geq 3$, and a binomial logistic definition when $T = 2$. Here, as can be seen in the denominator, we have selected type $t = 1$ as the reference category in the structural model. The coefficients for the reference type are set equal to 0, leading to $\exp(\boldsymbol{\gamma}_1'\mathbf{Z}) = \exp(\mathbf{0}'\mathbf{Z}) = \exp(0) = 1$. This selection of the first type, specifically, is computationally arbitrary. Different LCA software implement different arbitrary default selections of the reference type. For example, `Mplus` sets class T as reference, and `multilevLCA` sets class 1 as reference.

Because of the distinct conceptualizations of the measurement model for the relationship between the unobserved types and their observed indicators, and the structural model for the relationship between the unobserved types and their observed predictors, best practice is to estimate and analyze these two model components separately. This involves two fundamental practices. One practice is implicit in the previous section and its separation from this section: performing model selection for T prior to the inclusion of the covariates (e.g., Masyn, 2017). The second practice involves taking selection of T as given: estimating the structural model with the covariates by means of a *stepwise estimation approach* (Bakk & Kuha, 2018; Vermunt, 2010). This family of estimation approaches is used to avoid interpretational confounding of the identified types when the covariates are added to the model.

There are different approaches for fitting the structural model with covariates in Equation (3). We focus the exposition of the main estimation approaches on the concepts over technical detail. These approaches were developed to overcome the drawbacks of the *one-step* estimation approach in which the full parameter space for the full model in Equation (2) is fitted simultaneously. While the one-step approach is seemingly natural, it allows for interpretational confounding in that the inclusion of the covariates may distort the measurement model and the definition of the LCs. Furthermore, when the applied researcher is comparing structural models with different sets of covariates, the one-step approach is computationally inefficient, in that the measurement model is re-fitted every time changes are made to the structural model. Stepwise estimators overcome these defects by separating the estimation of the measurement model from the subsequent estimation of the structural model.

Recent literature indicates the advantages of using a stepwise estimator known as the *two-step* approach (Bakk & Kuha, 2018). This was developed as a more direct and more easily interpretable alternative to the *bias-adjusted three-step* approach (Vermunt, 2010). Both these approaches start with the same step 1: fitting the measurement model based on the standard LC model in Equation (1). The two-step

[3] In some analytical contexts, it is of interest to treat the types as predictors of an external target variable. We do not focus on such contexts in the present paper. Interested readers are referred to Masyn (2013) and Nylund-Gibson et al. (2019).

approach then fits in step 2 the structural model based on Equation (2) while fixing the $P(Y_h = r_h|X = t)$ at their fitted values from step 1. In contrast, step 2 of the bias-adjusted three-step approach involves assigning the units in the sample to the classes based on the fitted values for the measurement model from step 1. Then, in step 3, a model for the assigned classes given **Z** is fitted while adjusting for misclassification in step 2. The statistical performance of these approaches have been shown to be as good as identical when class separation is strong and the classes easily distinguished, whereas the two-step is less biased and more efficient when class separation is weak and the classes are more difficult to distinguish (Bakk & Kuha, 2018). The statistical advantages of the two-step approach in addition to the conceptual advantages of its simple interpretations make the two-step approach preferable to the alternative bias-adjusted three-step approach.

However, historically many applied researchers have used the *naive three-step* approach. This has the same step 2 as the bias-adjusted three-step approach, but, in step 3, involves regressing the assigned classes on **Z** as if they were observed. Although the naive three-step approach is known in the methodological literature to be severely biased (Bakk & Kuha, 2018; Vermunt, 2010), some applied LCA still use this estimation approach. An explanation for this may be that the two-step approaches has until recently not been automatically implementable in specialized LCA open-source software. At the time of writing, the only packages implementing the two-step approach in `R` and `Python` are `multilevLCA` (Lyrvall, Di Mari, et al., 2025) and `StepMix` (Morin et al., 2025), respectively. Nevertheless, the application of two-step estimation is growing with the availability of open-source software implementing the approach, particularly in `R`. For instance, in 2025 `multilevLCA` was used to apply two-step LCA and inform applied scholarship in health research (Najjuuko et al., 2025), transportation studies (Labbo et al., 2025), and economics (Angelini & Lyrvall, 2025). We provide a brief overview of specialized LCA software options in Section 5.

3. An Illustrative Empirical Example

We illustrate how the three fundamental research questions presented in the previous section can be answered by means of an empirical example. This example is based on real data from the European Social Survey (ESS), specifically ESS Round 11 (2023/24) (Sikt, 2025), for voting-age adults. For simplicity of exposition, we limit the geographical scope to 6 countries in and bordering South-Eastern Europe, namely, Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.[4] We focus on 8 survey items in the role of indicators, and 1 survey item in the role of covariate. The indicators are binary measures of participation in multiple political activities. The indicators include voting in the last national elections (75%), and participating in the last 12 months in a series of political activities: signing a petition (16%), posting or sharing anything about politics online (12%), volunteering for a nonprofit or charitable organization (10%), boycotting a certain product (10%), taking part in a public demonstration (8%), donating to or participating in a political party or pressure group (6%), and wearing or displaying a campaign badge or sticker (6%). The covariate measures the respondent's self-reported placement on the political left-right scale, from 0 for left to 10 for right (mode 5; median 5; mean 5.02; standard deviation 2.6). We include in the analysis only individuals who reported being eligible to vote at the time of the last national election. The total sample size for the applied analysis is 8078, ranging, by country, from 634 (Cyprus) to 2126 (Bulgaria). We conceptualize the (potential, at this stage of the analysis) underlying LC variable as types of political participants. The model estimation and graphical illustration is carried out in `R` using the `multilevLCA` package (Lyrvall, Di Mari, et al., 2025). Replication code is provided in Appendix C.1.

Table 3. Model selection criteria - the BIC, the AIC, the expected proportion of misclassification (P.miscl.), and the entropy-R^2 (R^2-ent.) - across 5 fitted candidate LC model specifications based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.

T	df	BIC	AIC	P.miscl.	R^2-ent.
1		44057.40	44001.43		
2	238	39331.45	39212.50	**0.05**	**0.73**
3	229	39187.47	39005.55	0.11	0.60
4	220	**39170.76**	38925.86	0.13	0.57
5	211	39206.01	**38898.15**	0.14	0.57

3.1. Do I Have Data with a Possible Clustering Structure That Can be Analyzed Using LCA?

Political participation can involve different repertoires across multiple different activities. People's observed repertoires of participation across these activities can be reasonably conceptualized as arising from a set of distinct ideal types of participants; or, people can be reasonably hypothesized to be clustered around a set of distinct ideal types on the basis of their participation across these activities. Such ideal types can be usefully described based on each type's proportion of the population and independent participation probabilities for each activity. The set of types is discrete, and the observed indicators used to identify it are discrete. As such, we argue that LCA is suitable for analyzing types of civic-political participants, using the 8 binary survey items in the role of indicators.

[4]This decision is driven by the illustrative value of providing the reader with a consistent empirical example across the present section and the subsequent section. More specifically, the data for this subset of countries yield a homogeneous measurement model across this example and two examples illustrating differential item functioning modeling and multilevel modeling. This is important, given the primarily illustrative aims of this paper, because we discuss these three different LC models as the final specification of their (sub)sections. In a study aiming to make a substantive contribution, there would reasonably be broader substantive value in including survey data on a larger number of theoretically relevant countries.

Table 4. Fitted three-class LC model based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.

	Class 1 ("activists")	Class 2 ("mainstream")	Class 3 ("vote specialist")
Proportion	0.045	0.241	0.714
Response probability			
vote	0.932	0.853	0.703
sign petition	0.883	0.455	0.014
post online	0.770	0.299	0.019
volunteer	0.558	0.230	0.033
boycott	0.492	0.256	0.025
demonstrate	0.760	0.162	0.004
donate	0.435	0.139	0.006
wear badge	0.493	0.132	0.004

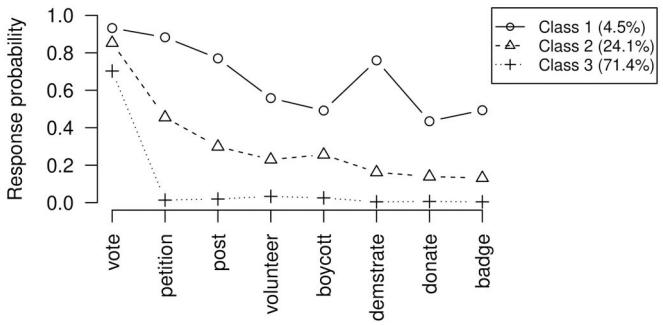


Figure 3. Fitted three-class LC model based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia; sample mean values are: 0.75 for *vote*, 0.16 for *petition*, 0.12 for *post*, 0.10 for *volunteer*, 0.10 for *boycott*, 0.08 for *demonstrate*, 0.06 for *donate*, and 0.06 for *badge*.

3.2. What Optimal Number of LCs Describe the Possible Clustering Structure in my Data?

Because there is no ex-ante substantive theory of the "correct" number of LCs, we perform model selection across a set of candidate specifications. As the smallest candidate specification, we consider the one-class model, to test if the data have any LC structure whatsoever. As the largest candidate specification, we consider the five-class model. All four considered LC models to be estimated (that is, the two-class model to the five-class model—excluding the one-class model because it does not involve estimation) have positive degrees of freedom and are thus statistically identified. We consider multiple selection criteria involving information criteria and measures of class separation. Specifically, we consider the BIC, the AIC, the expected proportion of misclassification, and the entropy-R^2. For the former three of these criteria, smaller quantities are preferable, and for the latter, larger quantities are preferable.

Table 3 presents the degrees of freedom and selection criteria for the fitted candidate specifications. For each selection criterion, the optimal specification is emboldened. The information criteria suggest different candidates as optimal. Because the candidates suggested by the BIC and the AIC do not include the one-class model, we can conclude that the data have an underlying LC structure. The BIC suggests the three-class model, and the AIC the four-class model. The expected proportion of misclassification and the entropy-R^2 suggest the two-class model.

Next, we consider the substantive interpretations of the models. We begin by interpreting the fitted three-class model, which is illustrated in Table 4 and Figure 3. From the measurement model, we can observe that class 1 has the highest participation probabilities for all 8 activities. The participation probabilities are smaller than 0.5 for boycotting a product, donating to or participating in a political party or pressure group, and wearing or displaying a campaign badge or sticker, yet substantially higher than for the other classes. Class 2 has a pattern that is broadly substantively consistent with the participation probabilities in the full sample, yet the participation probabilities of class 2 are consistently higher. Class 3 scores high on voting, and very low on the remaining 7 activities. From the fitted structural model, we can observe that class 1 is the least common, representing about 5% of people in the countries in focus; class 2 represents about 25%; and class 3 is the substantially most common pattern of participation, representing a majority of about 70%.

These class profiles are strikingly similar to previous political research by Oser (2017), who identified a LC model with 4 classes of political participation based on survey data from a different geographical and temporal context, namely, the United States in 2005. The three classes identified in the present empirical example are consistent with three of the classes in Oser's four-class solution, both with respect to the measurement model and the structural model. We label the identified classes in the present empirical example as in Oser (2017); that is, we label class 1 "activists," class 2 "mainstream," and class 3 "vote specialist."

The fitted two-class model (Figure B1 in Appendix B) is less substantively interesting. This model identifies the "vote specialists," and combines "activists" and "mainstream" into a single class. While this is a *quantitatively* more well-separated class solution (based on classification accuracy and entropy), *substantively* we consider the three-class solution with the separate "activist" and "mainstream" classes to be more distinct and meaningful. Furthermore, we consider the fitted four- and five-class models (Figure B2 and Figure B3 in Appendix B, respectively) to be less substantively interesting. Both models split "mainstream" into two new classes that are not distinguished in any substantively interesting manner. The five-class solution also splits "activist" into two new classes. One of them is rather similar to the original "activist" class, and the other is a corner solution with very low prevalence (1% of individuals) which can be taken to be an outlier pattern within the original "activist" class.

Taking these quantitative and substantive results into account, overall we select the three-class model for its locally optimal balance of fit, class separation, and interpretability.

3.3. Can the Identified LCs in my Data be Predicted Based on External Covariates?

We estimate the association between the identified types, as dependent variable, and left-right ideological political placement, as independent variable, or covariate. Political placement is here treated as a numerical variable (integer values from 0 to 10), where higher values correspond to more

Table 5. Fitted structural model for the three-class LC model, parameterized as multinomial regression with class 3 as reference category, based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.

	Class 2 ("mainstream")	Class 1 ("activists")
left-right scale	0.004 (0.014)	−0.131*** (0.027)
(*Bulgaria*)		
Croatia	−0.045 (0.123)	−1.932** (0.835)
Cyprus	0.592*** (0.156)	0.848*** (0.314)
Israel	0.631*** (0.143)	1.739*** (0.250)
Montenegro	0.542*** (0.119)	0.812*** (0.266)
Serbia	0.709*** (0.123)	1.159*** (0.248)
intercept	−1.303*** (0.138)	−2.625*** (0.271)

***$p < 0.01$.
**$p < 0.05$.
*$p < 0.1$.

right-leaning values. Country is also included as a covariate in the role of control variable. In the illustrative analysis, because we are not controlling for other variables like demographic and socio-economic characteristics, we do not aim to make an original contribution to political scholarship. The structural model with covariates is estimated by means of the two-step estimation approach (Bakk & Kuha, 2018). Due to missing values on the covariate, the sample size for the second estimation step for the regression is reduced to 6722. Because the measurement model is estimated separately in step 1, this loss of observations does not have any adverse effects on the identification of the class solution and the dependent variable in the regression. This illustrates one of the strengths of stepwise estimation.

We specify the multinomial logistic regression model with the majority "vote specialist" class as reference category. The two-step estimates for the structural model with covariates are reported in Table 5. Political placement is estimated to have no statistically significant role in distinguishing "mainstream" from "vote specialists." However, in distinguishing "activists" from "vote specialists," political placement is estimated to have a (highly) statistically significant role. The logit coefficient of −0.13 suggests that the odds of being an "activist" to being a "vote specialist" are lower for more right-leaning people than for more left-leaning people. On the basis of these results, we can conclude that, in the countries in focus, political orientation tends to distinguish "activists" from "vote specialists," such that "activists" has a larger proportion of left-leaning people. In contrast, the findings show no distinction in left-right orientation between "mainstream" and "vote-specialists."

4. Modeling Extensions for More Complex LCA

The LCA methodological literature has developed extensions of the standard modeling framework for analytical contexts in which some of its assumptions are violated. This section summarizes the two most common more complex LC models that are highly relevant for applied researchers: models with differential item functioning (DIF), and multilevel LC models.

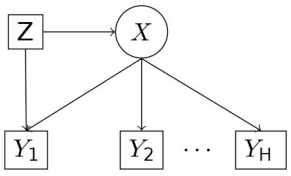


Figure 4. Path diagram for the LC model with DIF in one indicator (Y_1).

4.1. Differential Item Functioning

The concept of DIF is also known as direct effects, measurement invariance, and measurement nonequivalence. LC models with DIF include external confounding variables in the relationship(s) between X and one or more Y_h. As an illustrative example of DIF, consider survey research replicating the freedom of speech norms analysis from Sub-section 2.1 in a cross-national setting across the United States and Mexico. Say that, due to translation error, the survey item for allowing anti-religionists to speak is worded such that this policy is presented as substantially less severe in the Spanish-language survey than in the English-language survey. A Mexican respondent is thus more likely than an American respondent to respond in favor of the policy, beyond differences in their freedom of speech norms. In estimating the LC model, this external and systematic variation in response probabilities across survey languages disturbs the functioning of the survey item as a measurement of the underlying structure of norms, and the indicator has DIF on the basis of survey language. The path diagram in Figure 4 graphically illustrates a LC model with DIF for one indicator.

The extension of the LC model with covariates to the DIF modeling context is straightforward. The confounding variables giving rise to DIF may or may not also serve as covariates. Let $\mathbf{Z}^* = (1, Z_1^*, ..., Z_M^*)'$ be a set of M external variables giving rise to DIF, and a constant 1 to allow for an intercept. $\mathbf{Z}$ and $\mathbf{Z}^*$ may contain common variables (for instance, in the example of freedom of speech norms with DIF, survey language can serve both as a source of DIF and as a covariate). We can write the LC model with DIF as

$$P(Y_1 = r_1, ..., Y_H = r_H|\mathbf{Z}, \mathbf{Z}^*) = \sum_{t=1}^{T} P(X = t|\mathbf{Z}, \mathbf{Z}^*) \prod_{h=1}^{H} \prod_{r_h=1}^{R_h} P(Y_h = r_h|X = t, \mathbf{Z}^*)^{I(Y_h = r_h)},$$

where the structural model is defined in Equation (3) as in the standard modeling context in the absence of DIF. The measurement model with DIF is typically parameterized by means of logistic coefficients (e.g. Vermunt & Magidson, 2021a). We can write this as

$$P(Y_h = r_h|X = t, \mathbf{Z}^*) = \frac{\exp(\boldsymbol{\beta}'_{rht}\mathbf{Z}^*)}{1 + \sum_{q=2}^{R_h} \exp(\boldsymbol{\beta}'_{qht}\mathbf{Z}^*)},$$

where $\boldsymbol{\beta}_{rht} = (\beta_{rht0}, \beta_{rht1}, ..., \beta_{rhtM})'$ is the set logistic coefficients, consisting of an intercept β_{rht0} and M logistic slope coefficients corresponding to $Z_1^*, ..., Z_M^*$. These coefficients are defined as multinomial logistic definition when $R_h \geq 3$, and binomial logistic when $R_h = 2$. Here, response category

Table 6. Fitted three-class LC measurement model with DIF based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.

	Class 1 ("activists")	Class 2 ("mainstream")	Class 3 ("vote specialist")
Response probability			
vote	0.932	0.845	0.702
sign petition	0.872	0.431	0.011
post online (*internet user = 0*)	**0.409**	**0.137**	**0.001**
post online (*internet user = 1*)	**0.786**	**0.312**	**0.026**
volunteer	0.535	0.224	0.031
boycott	0.481	0.246	0.023
demonstrate	0.736	0.148	0.003
donate	0.415	0.135	0.004
wear badge	0.481	0.116	0.005

Table 7. Fitted three-class LC structural model with DIF based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.

	Class 2 ("mainstream")	Class 3 ("activists")
internet user	0.871	1.490
intercept	−1.679	−3.833

$r_h = 1$ is set as the (computationally arbitrary) reference category in the measurement model, with the corresponding coefficients set equal to 0. This general equation allows for indicators having no DIF, and DIF affecting a subset of the types t. For an indicator without DIF with respect to t, the slope coefficients in $\boldsymbol{\beta}_{rht}$ are equal to 0, so that $\exp(\boldsymbol{\beta}'_{rht}\mathbf{Z}^*) = \exp((\beta_{rht0}, \mathbf{0})'\mathbf{Z}^*) = \exp(\beta_{rht0})$.

Articles by Masyn (2017), Vermunt and Magidson (2021a), and Lyrvall, Kuha, et al. (2025) are the benchmark references in the LCA methodological literature regarding the concept and mechanics of DIF. The effect of ignoring DIF have been studied by different sets of authors in recent years. It has been consistently shown that ignoring DIF can yield substantively biased estimates of the covariate effects (Asparouhov & Muthén, 2014; Di Mari & Bakk, 2018; Janssen et al., 2019; Lyrvall, Kuha, et al., 2025; Masyn, 2017). The methodological literature has proposed extensions of stepwise estimation approaches for modeling DIF and reducing this bias. Such proposals have been put forth by Vermunt and Magidson (2021a) for the bias-adjusted three-step estimation approach, and by Lyrvall, Kuha, et al. (2025) for the two-step estimation approach in the multilevel context (we introduce the multilevel context in the next section). We here give a brief summary of the key idea shown by these authors. The key idea is to include $\mathbf{Z}^*$ in the estimation of the measurement model in the first step, thus fitting $P(X = t|\mathbf{Z}^*)$ and $P(Y_h = r_h|X = t, \mathbf{Z}^*)$ simultaneously. In the following estimation of the structural model, the step-1 fitted values for $P(X = t|\mathbf{Z}^*)$ are discarded, the step-1 fitted values for $P(Y_h = r_h|X = t, \mathbf{Z}^*)$ retained and held fixed, and the full parameter space for structural model $P(X = t|\mathbf{Z}, \mathbf{Z}^*)$ fitted to obtain the final stepwise estimates.

However, the question of how to identify DIF has not yet been settled. We consider the lack of a rigorously evaluated and generally accepted, detailed benchmark model selection approach for models with DIF as constituting a meaningful gap in the literature, opening an interesting avenue for future research. However, some implicit recommendations in the methodological literature are clear from its application in illustrative LCAs with DIF by Masyn (2017) and Lyrvall, Kuha, et al. (2025). The first recommendation is to select the number of types T with respect to the standard LC model in Equation (1). This is analogous to the general recommendation for LC models with covariates. When the number of types has been selected, different DIF specifications can be compared to determine what variables have DIF and if DIF varies by LC or not. The second recommendation is to base the selection of DIF specification on the same decision criteria that are used to select the number of types.

We now give an empirical example of DIF by extending the three-class LC model of political participation types in Table 4 and Figure 3. Specifically, we include internet usage as a covariate giving rise to DIF. Internet usage is observed on a 5-point scale in the original ESS data. For conciseness of illustration, we collapse the response categories and create a binary variable that is equal to 1 if the respondent uses the internet on most days or every day, and equal to 0 if the respondent uses the internet a few times a week, only occasionally, or never. While we dichotomized the covariate for illustrative purposes, in most applications dichotomization is not recommended. Also for conciseness of illustration, we do not perform model selection for the DIF specification, directly modeling non-uniform DIF in the indicator for posting or sharing anything about politics online. The substantive reason for this choice is that we want the political participation types to allow people with different internet use habits to have different propensities of online political activism. Note that, because we are considering a single covariate, in this example $\mathbf{Z}^* = \mathbf{Z}$. Replication code for this empirical example is provided in Appendix D.

The fitted measurement parameters are reported in Table 6, in which the response probabilities for posting or sharing about politics are emboldened. Beyond the DIF extension of the model, the measurement model is as good as identical to the class solution without DIF. With respect to DIF, we can observe that individuals with higher overall internet usage have substantially higher probabilities of posting or sharing about politics than individuals with lower overall internet usage. Table 7 reports the two-step point estimates. The logit coefficients of 0.87 and 1.49 suggest that people with higher internet usage have higher odds of being "mainstream" to being a "vote specialist", and even higher odds of being an "activist" to being a "vote specialist."

4.2. Multilevel LCA

Multilevel LC modeling is used for hierarchical data. In applied LCA, hierarchical data analyzed with multilevel LC modeling often take the form of cross-national survey data. In such cases, the units of primary analysis are the survey respondents, and the units of secondary analysis the nations in which the respondents are nested. For example, in their separate consumer segmentation studies using survey data, Bijmolt et al. (2004) analyze individuals nested within nations, and Paccagnella and Varriale (2013) analyze households nested within nations. A different example of

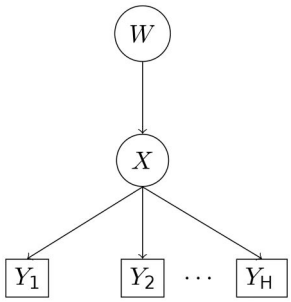


Figure 5. Path diagram for the multilevel LC model.

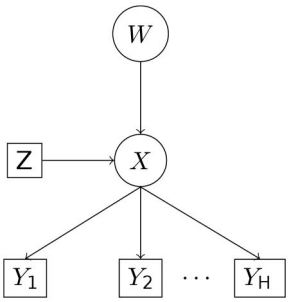


Figure 6. Path diagram for the multilevel LC model.

hierarchical data analyzed using multilevel LCA is students nested within classrooms (Fagginger Auer et al., 2016).

We focus our exposition of the multilevel modeling context on the commonly considered random-effect approach (Vermunt, 2003). The extension of Equation (1) to the multilevel modeling context involves operationalizing the structural model $P(X = t)$ as a mixture model on the basis of a second LC variable defined on the higher level of the hierarchical data. Let W be the higher-level LC variable, and M its number of categories $m = 1, ..., M$. Assuming that the responses are independent of W given X, as is standard in the multilevel LCA literature (Vermunt, 2003), we can write the multilevel LC model as

$$P(Y_1 = r_1, ..., Y_H = r_H) = \sum_{m=1}^{M} P(W = m) \sum_{t=1}^{T} P(X = t|W = m) \prod_{h=1}^{H} \prod_{r_h=1}^{R_h} P(Y_h = r_h|X = t)^{I(Y_h=r_h)}.$$

This is graphically illustrated in Figure 5 by means of a path diagram. The multilevel LC model can be described as a "mixture of mixtures": the measurement model is a mixture of the categories of the standard (single-level) LC variable X, and the structural model a mixture of the categories of the higher-level LC variable W. The multilevel structure can be easily interpreted, with respect to some individual i nested within some nation j, in the following way. First, the national-level type of j is realized on the basis of the marginal higher-level class probabilities $P(W = m)$. Second, given the higher-level type, the individual-level type of i is realized on the basis of the conditional lower-level class probabilities $P(X = t|W = m)$. Finally, the responses are realized on the basis of the lower-level type in the same way as in the single-level modeling context.

Model selection for multilevel LC models is typically performed based on the benchmark approach by Lukočienė et al. (2010). Consistent with the general recommendations for the single-level context, this involves selecting the number of LCs without any covariates. Their so-called "sequential" model selection approach involves the following three stages. First, the number of types is selected for the standard (single-level) LC model in Equation (1), ignoring the multilevel structure. Second, the selected number of types is held fixed while the number of higher-level LCs is selected for the multilevel LC model (by comparing a set of alternatives in an analogous way to model-selection in the single-level context). Finally, the selected number of higher-level LCs is held fixed while the number of types is selected again, this time for the multilevel model. The final specification is given by the number of higher-level LCs from the second stage and the number of types from the final stage.

Because of the presence of two units of analysis, the multilevel extension of Equation (2) can include covariates both on the lower level and on the higher level. Let $\mathbf{Z}^H = (1, Z_1^H, ..., Z_V^H)'$ be a set of V higher-level covariates and a constant 1 to allow for an intercept. For example, if the higher-level unit of analysis is the nation of the respondent, $\mathbf{Z}^H$ may include variables such as population, gross domestic product per capita, and risk premium on government bonds. Making the equivalent assumption from the single-level context that the indicators are conditionally independent of the higher-level covariates given the LC variable(s), we obtain

$$P(Y_1 = r_1, ..., Y_H = r_H|\mathbf{Z}, \mathbf{Z}^H) = \sum_{m=1}^{M} P(W = m|\mathbf{Z}^H) \sum_{t=1}^{T} P(X = t|W = m, \mathbf{Z}) \prod_{h=1}^{H} \prod_{r_h=1}^{R_h} P(Y_h = r_h|X = t)^{I(Y_h=r_h)}.$$

Figure 6 graphically illustrates a multilevel LC model with covariates by means of a path diagram. We can accordingly extend Equation (3) to

$$P(X = t|W = m, \mathbf{Z}) = \frac{\exp(\boldsymbol{\gamma}'_{t|m}\mathbf{Z})}{1 + \sum_{s=2}^{T} \exp(\boldsymbol{\gamma}'_{s|m}\mathbf{Z})}.$$

This specification allows for different covariate effects on the same lower-level LC across the higher-level LCs. The higher-level structural model can be specified as

$$P(W = m|\mathbf{Z}^H) = \frac{\exp(\boldsymbol{\alpha}'_m\mathbf{Z}^H)}{1 + \sum_{l=2}^{M} \exp(\boldsymbol{\alpha}'_l\mathbf{Z}^H)},$$

where $\boldsymbol{\alpha}_m = (\alpha_{m0}, \alpha_{m1}, ..., \alpha_{mV})'$ is the set of higher-level logistic coefficients - specifically, multinomial logistic when $M \geq 3$, and binomial logistic when $M = 2$ - consisting of an intercept α_{m0} and V logistic slope coefficients corresponding to $Z_1^H, ..., Z_V^H$. Response category $m = 1$ is set as the (computationally arbitrary) reference category, with 0-valued corresponding coefficients.

We recommend the paper by Vermunt (2003) as the benchmark methodological reference regarding the concepts and mechanics of multilevel LC modeling. Stepwise estimation approaches for multilevel models with covariates have recently been proposed in the LCA methodological literature. Lyrvall et al. (2024) extend the bias-adjusted three-step

Table 8. Fitted structural model for the multilevel LC model with 3 lower-level (LL) LCs and 2 higher-level (HL) LCs, based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia.

	HL Class 1 ("more engaged")	HL Class 2 ("less engaged")
HL proportion	0.667	0.333
LL proportion		
LL Class 1 ("activists")	0.079	0.027
LL Class 2 ("mainstream")	0.322	0.205
LL Class 3 ("vote specialist")	0.599	0.768

approach to the multilevel modeling context. Di Mari et al. (2023) extend the two-step approach to the multilevel modeling context. Lyrvall, Kuha, et al. (2025) extend the two-step approach to the multilevel modeling context with DIF.

We concisely illustrate a multilevel LC model by means of an empirical example, extending the fitted single-level LC measurement model in Table 4. In applied research, multilevel modeling is typically used with more than 6 higher-level units. The decision of using this subset of the data with 6 countries is driven by illustrative purposes; see Footnote 1. In a study aiming to make a substantive contribution, there would reasonably be broader substantive value in including survey data on a larger number of theoretically relevant countries. As higher-level unit of analysis, we use the country of the respondent. The conceptual justification for modeling cross-country variation by means of multilevel modeling, as opposed to including country as a covariate like in the structural model in Table 5, is the substantive interest and interpretational simplicity in identifying distinct clusters of countries with respect to their proportions of different political participation types, compared to analyzing these proportions in each individual country. We illustrate multilevel modeling by including 2 higher-level classes. The model estimation is carried out in the `R` package `multilevLCA`. Replication code is provided in Appendix C.2.

For brevity, we present only the structural models on the higher level and the lower level for the fitted model. The fitted structural parameters are presented in Table 8. As can be seen, in higher-level class 1, the "activists" and the "mainstream" are substantially more prevalent compared to higher-level class 2. The "vote-specialists" are the majority type in both higher-level classes, but distinctly more prevalent in higher-level class 2. As such, we can label higher-level class 1 "more engaged" and higher-level class 2 "less engaged." About 67% of the countries in focus belong to the "more engaged" class, and 33% to the "less engaged" class. In applied research, when the higher-level units are of substantive interest, their assigned higher-level classes are often reported. Class assignments are typically provided by LCA software.[5] In the present analysis, we report that Bulgaria and Croatia are assigned to the "less engaged" class, and Cyprus, Israel, Montenegro, and Serbia assigned to the "more engaged" class.

[5]For formulae and broader methodological contexts of class assignment, we refer interested readers to Vermunt (2010) for the standard, single-level context, Vermunt and Magidson (2021a) for the DIF modeling context, and Vermunt (2003) for the multilevel modeling context.

5. Software

How can the discussed models be estimated in practice? Among proprietary software options, we consider `LatentGOLD` (e.g., Vermunt & Magidson, 2021b) and `Mplus` (e.g., Muthén & Muthén, 2019) to be the benchmark options, implementing all the discussed models. Both of them have been widely used in the applied LCA literature. For example, `LatentGOLD` was used in political research by Oser (2022) for standard LC modeling, in health research by Maas et al. (2018) for LC modeling with DIF (in model selection, finding no evidence of DIF), and in educational research by Fagginger Auer et al. (2016) for multilevel LC modeling. For instance, `Mplus` was used in substance abuse research by Bray et al. (2023) for standard LC modeling, in health research by Bettencourt et al. (2022) for LC modeling with DIF, and in educational research by Allison et al. (2016) for multilevel LC modeling.

Among open-source software options, we consider the benchmark option in `Python` to be the package `StepMix` (Morin et al., 2025), and the benchmark option in `R` to be the package `multilevLCA` (Lyrvall, Di Mari, et al., 2025), because they offer the largest set of functionalities in their respective environments. In particular, these are currently the only packages estimating LC models with covariates using stepwise approaches, including the two-step approach (Bakk & Kuha, 2018). However, neither of these packages currently estimate LC models with DIF, and `StepMix` does not estimate multilevel LC models. For example, `StepMix` was used in health research by Ng et al. (2025) for standard LC modeling. For instance, `multilevLCA` was used in economic research by Angelini and Lyrvall (2025) for standard LC modeling, and in health research by Najjuuko et al. (2025) for multilevel LC modeling.

`R` offers some alternative specialized LCA software. The `poLCA` (Linzer & Lewis, 2011) package was developed more than a decade before `multilevLCA` and is still widely used. The package estimates single-level LC models. The more recent `glca` package (Kim et al., 2022) estimates single-level and multilevel LC models when all the indicators are binary. In contrast, `poLCA` and `multilevLCA` allow for binary and polytomous indicators. Both `poLCA` and `glca` allow for the inclusion of covariates, and estimate such models using only the one-step approach. For example, `poLCA` was used in political research by Bertsou and Caramani (2022), and `glca` was used in environmental research by Aguión et al. (2025). On the other hand, `multilevLCA` uses the preferable option of giving researchers the option to estimate either the one-step or two-step approach.

We summarize these functionalities in Table 9.

6. Concluding Remarks

This paper provided an overview of statistical frameworks and best practices for the application of latent class analysis. The exposition focused on three fundamental research questions in applied LCA, in the order they would typically be answered: (1) *do I have data with a possible clustering structure that can be analyzed using LCA?*; (2) *what optimal*

Table 9. Functionalities of six specialized LCA software. All the listed options estimate single-level models for binary indicators.

Software	Access	Polytomous indicators	Stepwise estimation	Multilevel modeling	DIF modeling
`LatentGOLD`	Proprietary	√	√	√	√
`Mplus`	Proprietary	√	√	√	√
`StepMix` (`Python`)	Open source	√	√		
`multilevLCA` (`R`)	Open source	√	√	√	
`poLCA` (`R`)	Open source	√			
`glca` (`R`)	Open source			√	

number of LCs describe the possible clustering structure in my data?; and (3) *can the identified LCs in my data be predicted based on external covariates?*. We take the basic modeling choice of LCA as given, omitting an implied "research question" 0 of whether this categorical latent variable approach is more suitable than alternative approaches like linear scales.

The main methodological focus was on single-level LC models with covariates. The more complex methodological frameworks of differential item functioning and multilevel modeling were presented more briefly. The discussed statistical frameworks and best practices were illustrated by means of empirical examples, with replication code provided in the appendices of this paper. Where the size and scope of this paper were insufficient for detailed treatments of certain topics, the reader was referred to external benchmark articles.

An overview of software options for the estimation of LC models was provided. More specifically, we discussed the proprietary sofware `LatentGOLD` and `Mplus` and open-source software, including the `Python` package `StepMix`, and the `R` packages `multilevLCA`, `poLCA`, and `glca`. Because most specialized LCA open-source software are available in `R`, most of the discussed software are packages in that environment.

This paper is aimed at serving as a teaching resource supporting researchers in meaningfully applying LCA. It does so by clarifying the technical skill and theoretical judgment that is required at several steps of application.

Acknowledgment

We thank Jeremy Albright, Anna Lia Brunetti, Christopher Hare, Oliver Lang, Jonathan N. Katz, Linette Lim, Amanda Weiss, and Barak Zur for helpful comments, as well as participants at presentations at the Midwest Political Science Association conference in April 2026, and the Visions in Methodology conference in May 2026.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Research Council under Grant 101077659.

ORCID

Johan Lyrvall http://orcid.org/0000-0002-1863-8147
Jennifer Oser http://orcid.org/0000-0002-1531-4606
Zsuzsa Bakk http://orcid.org/0000-0001-9352-4812

References

Aguión, A., Basurto, X., Funge-Smith, S., Gorelli, G., Iversen, E., Mancha-Cisneros, M., & Gutierrez, N. (2025). Five archetypes of small-scale fisheries reveal a continuum of production strategies to guide governance and policymaking. *Nature Food*, *6*, 1020–1031. https://doi.org/10.1038/s43016-025-01237-5

Allison, K., Adlaf, E., Irving, H., Schoueri-Mychasiw, N., & Rehm, J. (2016). The search for healthy schools: A multilevel latent class analysis of schools and their students. *Preventive Medicine Reports*, *4*, 331–337. https://doi.org/10.1016/j.pmedr.2016.06.016

Angelini, F., & Lyrvall, J. (2025). Art on trial: Mapping public reactions to scandalous cultural goods in public institutions. *Applied Economics Letters*. Advance online publication. https://doi.org/10.1080/13504851.2025.2549507

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M*plus*. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 329–341. https://doi.org/10.1080/10705511.2014.915181

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, *83*, 871–892. https://doi.org/10.1007/s11336-017-9592-7

Bakk, Z., & Kuha, J. (2021). Relating latent class membership to external variables: An overview. *The British Journal of Mathematical and Statistical Psychology*, *74*, 340–362. https://doi.org/10.1111/bmsp.12227

Bertsou, E., & Caramani, D. (2022). People haven't had enough of experts: Technocratic attitudes among citizens in nine European democracies. *American Journal of Political Science*, *66*, 5–23. https://doi.org/10.1111/ajps.12554

Bettencourt, A., Musci, R., Masyn, K., & Farrell, A. (2022). Latent classes of aggression and peer victimization: Measurement invariance and differential item functioning across sex, race-ethnicity, cohort, and study site. *Child Development*, *93*, e117–e134. https://doi.org/10.1111/cdev.13691

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 719–725. https://doi.org/10.1109/34.865189

Biernacki, C., Celeux, G., & Govaert, G. (2010). Exact and Monte Carlo calculations of integrated likelihoods for the latent class model. *Journal of Statistical Planning and Inference*, *140*, 2991–3002. https://doi.org/10.1016/j.jspi.2010.03.042

Bijmolt, T., Paas, L., & Vermunt, J. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, *21*, 323–340. https://doi.org/10.1016/j.ijresmar.2004.06.002

Bozdogan, H. (1993). Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In *Information and Classification: Concepts, Methods and Applications Proceedings of the 16th Annual Conference of the "Gesellschaft für Klassifikation ev" University of Dortmund*, April 1–3, 1992 (pp. 40––54.).

Bray, B. C., Watson, D. P., Salisbury-Afshar, E., Taylor, L., & McGuire, A. (2023). Patterns of opioid use behaviors among patients seen in the emergency department: Latent class analysis of baseline data from the POINT pragmatic trial. *Journal of Substance Use and Addiction Treatment*, *146*, 208979. https://doi.org/10.1016/j.josat.2023.208979

Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. John Wiley & Sons.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *39*, 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Di Mari, R., & Bakk, Z. (2018). Mostly harmless direct effects: A comparison of different latent Markov modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 467–483. https://doi.org/10.1080/10705511.2017.1387860

Di Mari, R., Bakk, Z., Oser, J., & Kuha, J. (2023). A two-step estimator for multilevel latent class analysis with covariates. *Psychometrika*, *88*, 1144–1170. https://doi.org/10.1007/s11336-023-09929-2

Dziak, J., Lanza, S., & Tan, X. (2014). Effect size, statistical power, and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. *Structural Equation Modeling: a Multidisciplinary Journal*, *21*, 534–552. https://doi.org/10.1080/10705511.2014.919819

Fagginger Auer, M., Hickendorff, M., Van Putten, C., Béguin, A., & Heiser, W. (2016). Multi-level latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, *29*, 144–159. https://doi.org/10.1080/08957347.2016.1138959

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis*. Cambridge University Press.

Janssen, J., Van Laar, S., De Rooij, M., Kuha, J., & Bakk, Z. (2019). The detection and modeling of direct effects in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 280–290. https://doi.org/10.1080/10705511.2018.1541745

Kim, M., Vermunt, J., Bakk, Z., Jaki, T., & Van Horn, M. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 601–614. https://doi.org/10.1080/10705511.2016.1158655

Kim, Y., Jeon, S., Chang, C., & Chung, H. (2022). glca: An R package for multiple-group latent class analysis. *Applied Psychological Measurement*, *46*, 439–441. https://doi.org/10.1177/01466216221084197

Labbo, M., Jiang, X., Wada, S., de Dieu, G., & Bala, N. (2025). Road rage patterns in Nigeria: A multi-level latent class analysis. *Case Studies on Transport Policy*, *20*, 101471. https://doi.org/10.1016/j.cstp.2025.101471

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin.

Linzer, D., & Lewis, J. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*, 1–29. https://doi.org/10.18637/jss.v042.i10

Lukočienė, O., Varriale, R., & Vermunt, J. (2010). The simultaneous decision(s) about the number of lower-and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, *40*, 247–283. https://doi.org/10.1111/j.1467-9531.2010.01231.x

Lyrvall, J., Bakk, Z., Oser, J., & Di Mari, R. (2024). Bias-adjusted three-step multilevel latent class modeling with covariates. *Structural Equation Modeling: a Multidisciplinary Journal*, *31*, 592–603. https://doi.org/10.1080/10705511.2023.2300087

Lyrvall, J., Di Mari, R., Bakk, Z., Oser, J., & Kuha, J. (2025). Multilevel latent class analysis: State-of-the-art methodologies and their implementation in the r package multilevlca. *Multivariate Behavioral Research*, *60*, 731–747. https://doi.org/10.1080/00273171.2025.2473935

Lyrvall, J., Kuha, J., & Oser, J. (2025). Two-step multilevel latent class analysis in the presence of measurement non-equivalence. *Structural Equation Modeling: a Multidisciplinary Journal*, *32*, 678–687. https://doi.org/10.1080/10705511.2025.2490946

Maas, M., Bray, B., & Noll, J. (2018). A latent class analysis of online sexual experiences and offline sexual behaviors among female adolescents. *Journal of Research on Adolescence: The Official Journal of the Society for Research on Adolescence*, *28*, 731–747. https://doi.org/10.1111/jora.12364

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In *The oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis*. Oxford University Press.

Masyn, K. E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 180–197. https://doi.org/10.1080/10705511.2016.1254049

McCutcheon, A. L. (1987). *Latent class analysis*. (Vol. 64). Sage.

Moore, E. W. G., Quartiroli, A., & Little, T. D. (2025). Introduction to the best practice recommendations for longitudinal latent transition analysis. *International Journal of Psychology*, *60*, e70021. https://doi.org/10.1002/ijop.70021

Morin, S., Legault, R., Laliberté, F., Bakk, Z., Giguère, C.-É., de la Sablonnière, R., & Lacourse, É. (2025). Stepmix: A Python package for pseudo-likelihood estimation of generalized mixture models with external variables. *Journal of Statistical Software*, *113*, 1–39. https://doi.org/10.18637/jss.v113.i08

Muthén, B., & Muthén, L. (2019). Mplus: A general latent variable modeling program. *Muthén & Muthén*.

Najjuuko, C., Xu, Z., Kizito, S., Lu, C., & Ssewamala, F. (2025). Patterns of maternal and child health services utilization and associated socioeconomic disparities in Sub-Saharan Africa. *Nature Communications*, *16*, 7840. https://doi.org/10.1038/s41467-025-61350-8

Ng, H., Woodman, R., Veronese, N., Pilotto, A., & Mangoni, A. (2025). Comorbidity patterns and mortality in atrial fibrillation: A latent class analysis of the European study of older subjects with atrial fibrillation (eurosaf). *Annals of Medicine*, *57*, 2454330. https://doi.org/10.1080/07853890.2025.2454330

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*, 440–461. https://doi.org/10.1037/tps0000176

Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 967–985. https://doi.org/10.1080/10705511.2019.1590146

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In *Modern statistical methods for hci*. (pp. 275–287). Springer. https://doi.org/10.1007/978-3-319-26633-6_12

Oser, J. (2017). Assessing how participators combine acts in their "political tool kits": A person-centered measurement approach for analyzing citizen participation. *Social Indicators Research*, *133*, 235–258. https://doi.org/10.1007/s11205-016-1364-8

Oser, J. (2022). Protest as one political act in individuals' participation repertoires: Latent class analysis and political participant types. *American Behavioral Scientist*, *66*, 510–532. https://doi.org/10.1177/00027642211021633

Oser, J., Hooghe, M., Bakk, Z., & Di Mari, R. (2023). Changing citizenship norms among adolescents, 1999-2009-2016: A two-step latent class approach with measurement equivalence testing. *Quality & Quantity*, *57*, 4915–4933. https://doi.org/10.1007/s11135-022-01585-5

Oser, J., Hooghe, M., & Marien, S. (2013). Is online participation distinct from offline participation? a latent class analysis of participation types and their stratification. *Political Research Quarterly*, *66*, 91–101. https://doi.org/10.1177/1065912912436695

Paccagnella, O., & Varriale, R. (2013). Asset ownership of the elderly across europe: A multilevel latent class analysis to segment countries and households. In *Advances in theoretical and applied statistics*. (pp. 383–393). https://doi.org/10.1007/978-3-642-35588-2_35

Sikt. (2025). *European Social Survey European Research Infrastructure (ESS ERIC) (2025) ESS11 - integrated file, edition 3.0 [Data set]. Norwegian Agency for Shared Services in Education and Research*. https://doi.org/10.21338/ess11e030

Steinley, D. (2023). Mixture models. In *Handbook of structural equation modeling. :* Second edition. Guilford Press.

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*, 63–79. https://doi.org/10.1037/a0022673

Van Lissa, C., Garnier-Villarreal, M., & Anadria, D. (2024). Recommended practices in latent class analysis using the open-source R-package tidySEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 526–534. https://doi.org/10.1080/10705511.2023.2250920

Vermunt, J. (2003). Multilevel latent class models. *Sociological Methodology*, *33*, 213–239. https://doi.org/10.1111/j.0081-1750.2003.t01-1-00131.x

Vermunt, J. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450–469. https://doi.org/10.1093/pan/mpq025

Vermunt, J. (2025). Stepwise estimation of latent variable models: An overview of approaches. *Statistical Modelling*, *25*, 530–551. https://doi.org/10.1177/1471082X251355693

Vermunt, J., & Magidson, J. (2021a). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 356–364. https://doi.org/10.1080/10705511.2020.1818084

Vermunt, J., & Magidson, J. (2021b). *LG-syntax user's guide: Manual for LatentGOLD syntax module version 6.0*. Statistical Innovations Inc.

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: A guide to best practice. *Journal of Black Psychology*, *46*, 287–311. https://doi.org/10.1177/0095798420930932

White, A., & Murphy, T. (2014). BayesLCA: An R package for Bayesian latent class analysis. *Journal of Statistical Software*, *61*, 1–28. https://doi.org/10.18637/jss.v061.i13

Appendix A

Computation of a Standard Latent Class Model's Degrees of Freedom

The maximum number of types that can be statistically identified depends on the data. A specification is statistically identifiable if it has a non-negative number of degrees of freedom. This quantity is given by a combination of the number of possible unique response patterns in the data and the specification's number of freely estimable parameters. Let $\nu = \prod_{h=1}^{H} R_h$ be the number of possible unique response patterns in the data. Let ψ denote the specification's number of freely estimable parameters. For the structural model, $T-1$ parameters are freely estimable, and the remaining parameter $P(X=s)$ set to $1-\sum_{t\neq s} P(X=t)$ so that the class proportions sum to 1. For the measurement model, for each type t and indicator h, there are R_h-1 freely estimable parameters, and the remaining parameter $P(Y_h=q_h|X=t)$ is set to $1-\sum_{r\neq q} P(Y_h=r_h|X=t)$ so that the probabilities of the response categories sum to 1. Thus, $\psi = [T-1] + \left[T * \sum_{h=1}^{H}(R_h-1)\right]$. The degrees of freedom for the specification given the data is equal to $\nu - \psi - 1$.

To illustrate the computation of degrees of freedom in an applied context, we consider the freedom of speech norms analysis from the previous section. As shown in Table 1, the number of (possible) unique response patterns is $\nu = 8$. The number of freely estimable parameters for the structural model is equal to $T-1 = 2-1 = 1$. The number of freely estimable parameters for the measurement model is equal to $T * \sum_{h=1}^{H}(R_h-1) = 2 * \sum_{1}^{3}(2-1) = 6$. Thus, $\psi = [T-1] + \left[T * \sum_{h=1}^{H}(R_h-1)\right] = 1+6 = 7$. Thus further, the degrees of freedom for the two-class specification given the data is equal to $\nu - \psi - 1 = 8-7-1 = 0$. How many degrees of freedom would the hypothetical three-class specification have? Changing T from 2 to 3 does not affect the number of unique response patterns, but it does increase the number of freely estimable parameters from 7 to $[T-1] + \left[T * \sum_{h=1}^{H}(R_h-1)\right] = [3-1] + \left[3 * \sum_{1}^{3}(2-1)\right] = 11$. As such, the hypothetical three-class specification would have $\nu - \psi - 1 = 8-11-1 = (-4)$ degrees of freedom. Because this number is negative, the specification is not identifiable. As such, the data for the example can identify of up to 2 LCs.

Appendix B

Profile Plots for Selected Models in the Empirical Example in Section 3

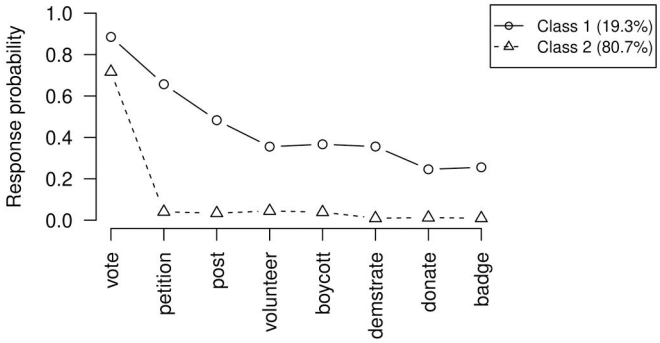


Figure B1. Fitted two-class LC model based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia; sample mean values are: 0.75 for *vote*, 0.16 for *petition*, 0.12 for *post*, 0.10 for *volunteer*, 0.10 for *boycott*, 0.08 for *demonstrate*, 0.06 for *donate*, and 0.06 for *badge*.

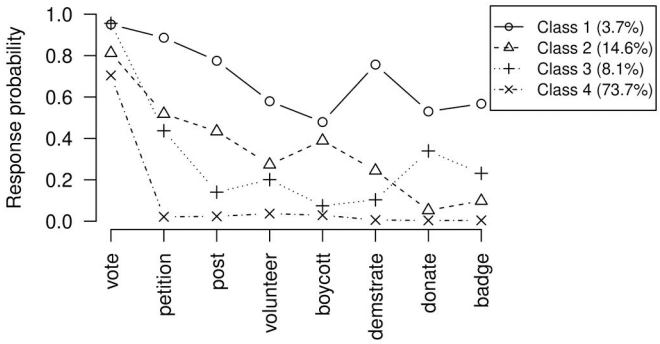


Figure B2. Fitted four-class LC model based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia; sample mean values are: 0.75 for *vote*, 0.16 for *petition*, 0.12 for *post*, 0.10 for *volunteer*, 0.10 for *boycott*, 0.08 for *demonstrate*, 0.06 for *donate*, and 0.06 for *badge*.

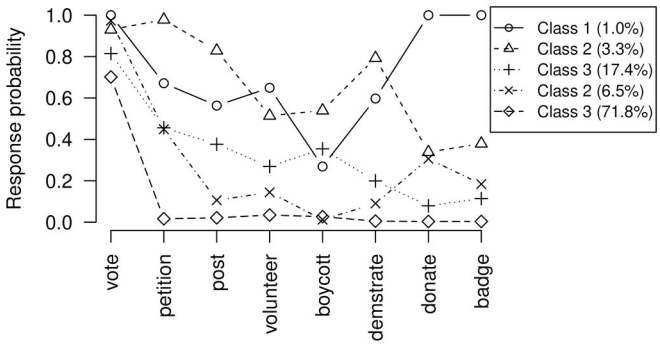


Figure B3. Fitted five-class LC model based on data from ESS Round 11 (2023/24) for Bulgaria, Croatia, Cyprus, Israel, Montenegro, and Serbia; sample mean values are: 0.75 for *vote*, 0.16 for *petition*, 0.12 for *post*, 0.10 for *volunteer*, 0.10 for *boycott*, 0.08 for *demonstrate*, 0.06 for *donate*, and 0.06 for *badge*.

Appendix C

Replication Code for the R Package multilevLCA

For brevity of code, we define the following three objects, which will be used in all model estimation with multilevLCA in this appendix. Here, ESS11 is the dataframe containing the observed data, and the vector with elements "vote" to "badge" contains the column names in ESS11 with the observed indicators.

```
data = ESS11
Y = c("vote", "petition", "post", "volunteer", "boycott",
      "demstrate", "donate", "badge")
iT = 3
```

The setting incomplete = TRUE is used in both sub-sections. This specifies that rows with missing values on the indicators should be retained in the estimation, given that at least one indicator value is observed, and the model estimated by means of full-information maximum likelihood. Setting incomplete = FALSE, which is the default and thus equivalent to ignoring the incomplete argument in the multiLCA call, removes rows with at least one missing value on any indicator before estimating the model (Lyrvall, Di Mari, et al., 2025).

For further brevity of code, we omit providing seed setting code, instead mentioning that set.seed(27) should be executed before each multiLCA call for exact replication.

C.1. The Illustrative Empirical Example in Section 3

The selected three-class model can be estimated by means of the script below. The first line stores the results in the object out, and the second line prints these results. This prints the BIC, the AIC, the expected proportion of misclassification, and the entropy-R^2 reported in Table 3, and the class proportions and the conditional response probabilities reported in Table 4.

```
out = multiLCA(data, Y iT, incomplete = TRUE)
out
```

Below, out is input to the multilevLCA plotting function, and the class labels manually written to include the class proportions (the default settings do not write the class proportions).

```
plot(out, clab = c("Class 1 (4.5%)", "Class 2 (24.1%)", "Class 3 (71.4%)"))
```

The two-step estimates reported in Table 5 can be replicated. This prints the results immediately, without saving the results in any object.

```
multiLCA(data, Y, iT, Z = c("left_right_scale", "country"),
          incomplete = TRUE, reord_user = c(3,2,1))
```

C.2. The Illustrative Empirical Example of Multilevel LCA in Section 4.2

The multilevel model can be estimated by means of the script below. This involves storing the results in out in the first line, and printing them in the second line. The multiLCA call involves specifying extout = TRUE for obtaining extensive output.

```
out = multiLCA(data, Y, iT, id_high = "country", iM = 2,
               incomplete = TRUE, extout = TRUE)
out
```

The call below prints the posterior higher-level class membership probabilities, which were output in the call above due to the setting extout = TRUE, rounded to two decimal points. This shows that Cyprus, Israel, Montenegro, and Serbia are assigned to the first higher-level class, and Bulgaria and Croatia are assigned to the second higher-level class.

```
round(out$mPW,2)
```

Appendix D

Replication Code for LatentGOLD

D.1. Step 1

```
options
  maxthreads = 8;
  algorithm
    tolerance = 1e-08 emtolerance = 0,01 emiterations = 250 nriterations = 50 ;
  startvalues
    seed = 0 sets = 16 tolerance = 1e-05 iterations = 50;
  bayes
    categorical = 1 variances = 1 latent = 1 poisson = 1;
```

```
  montecarlo
    seed = 0 sets = 0 replicates = 500 tolerance = 1e-08;
  quadrature nodes = 10;
  missing includedependent;
  output
    parameters = first betaopts = wl standarderrors profile probmeans = posterior
    loadings bivariateresiduals estimatedvalues = model reorderclasses;
variables
  dependent vote nominal, petition nominal, post nominal, volunteer nominal,
    boycott nominal, demstrate nominal, donate nominal, badge nominal;
  independent internet_user nominal;
  latent
    Cluster nominal 3;
equations
  Cluster <- 1 + internet_user;
  vote <- 1 + Cluster;
  petition <- 1 + Cluster;
  post <- 1 + Cluster + internet_user|Cluster;
  volunteer <- 1 + Cluster;
  boycott <- 1 + Cluster;
  demstrate <- 1 + Cluster;
  donate <- 1 + Cluster;
  badge <- 1 + Cluster;
```

D.2. Step 2

```
options
  maxthreads = 8;
  algorithm
    tolerance = 1e-08 emtolerance = 0,01 emiterations = 250 nriterations = 50 ;
  startvalues
    seed = 0 sets = 16 tolerance = 1e-05 iterations = 50;
  bayes
    categorical = 1 variances = 1 latent = 1 poisson = 1;
  montecarlo
    seed = 0 sets = 0 replicates = 500 tolerance = 1e-08;
  quadrature nodes = 10;
  missing includedependent;
  output
    parameters = first betaopts = wl standarderrors profile probmeans = posterior
    loadings bivariateresiduals estimatedvalues = model reorderclasses;
variables
  dependent vote nominal, petition nominal, post nominal, volunteer nominal,
    boycott nominal, demstrate nominal, donate nominal, badge nominal;
  independent internet_user nominal;
  latent
  Cluster nominal 3;
equations
  Cluster <- 1 + internet_user;
  vote <- (a1)1 + (a2)Cluster;
  petition <- (b1)1 + (b2)Cluster;
  post <- (c1)1 + (c2)Cluster + (c3)internet_user|Cluster;
  volunteer <- (d1)1 + (d2)Cluster;
  boycott <- (e1)1 + (e2)Cluster;
  demstrate <- (f1)1 + (f2)Cluster;
  donate <- (g1)1 + (g2)Cluster;
  badge <- (h1)1 + (h2)Cluster;
a1 = {
  0,8558
```

```
};
a2 = {
   0,8434
   1,7603
};
b1 = {
   -4,5019
};
b2 = {
   4,2223
   6,4236
};
c1 = {
   -6,7401
};
c2 = {
   4,9017
   6,3733
};
c3 = {
   3,1148
   1,0456
   1,6669
};
d1 = {
   -3,4465
};
d2 = {
   2,2042
   3,5870
};
e1 = {
   -3,7465
};
e2 = {
   2,6267
   3,6707
};
f1 = {
   -5,6749
};
f2 = {
   3,9260
   6,7000
};
g1 = {
   -5,5051
};
g2 = {
   3,6471
   5,1635
};
h1 = {
   -5,2083
};
h2 = {
   3,1729
   5,1338
};
```